SELF-ORGANIZING MAPS-BASED METHOD FOR THE ANALYSIS OF HEMARTHROSIS DATA IN HAEMOPHILIC PATIENTS

SOFÍA PÉREZ-ALENDA[1,2], JUAN J. CARRASCO[1,3], MARTA AGUILAR-RODRÍGUEZ[1,4], EMILIO SORIA-OLIVAS[3], JOSE MARÍA MARTÍNEZ-MARTÍNEZ[3], SANTIAGO BONANAD-BOIX[2], FELIPE QUEROL[1,2]

[1]Department of Physiotherapy, University of Valencia, C/ Gasc Oliag, 5, 46010, Valencia, Spain.

[2]Haemostasis and Thrombosis Unit, University and Polytechnic Hospital La Fe, Avda de Fernando Abril Martorell 106, 46026, Valencia, Spain.

[3]Intelligent Data Analysis Laboratory, E.T.S.E, University of Valencia, Avda Universitat S/N, 46100, Burjassot, Valencia, Spain.

[4]Rehabilitation Service, University and Polytechnic Hospital La Fe, Avda de Fernando Abril Martorell 106, 46026, Valencia, Spain.

Correspondence: Dra. Sofía Pérez-Alenda,

Departamento de Fisioterapia, Universitat de València,

C/ Gasco Oliag 5, 46010. Valencia, España.

Tel. +349639 83853, Fax +349639 83852, E-Mail: Sofia.Perez-Alenda@uv.es

Abstract: Joint bleeding are the most common injuries in hemophilia and if not treated properly can lead to disabling arthropathies from an early age. In order to ensure a good quality of life related to health in hemophilic patients is necessary a continuous monitoring of these injuries in addition to treatments that involve a very high cost to the health system. In this work, a database of hemofilia patients is analyzed using Self-Organizing Maps (SOM). All variables are depicted together, so that it

is possible to obtain the relationships between them, identifying possible clusters in data, and infer various characteristics of the tested population. Results obtained show that it is possible to obtain all this information in an intuitive and visual way, facilitating to the specialists the task of analysis.

Keywords: Hemarthrosis; Hemophilia; Ultrasonography; Visual Information Extraction; Self-Organizing Maps; Clinical Evaluation.

1. Introduction

Hemophilia is a hereditary disease linked to recessive X chromosome caused by a coagulation factor deficiency, the VIII in the case of hemophilia A and IX in hemophilia B (Mannucci & Tuddenham, 2001). It is classified according to the percentage in the blood of the deficient factor, being severe hemophilia when there is less than 1% factor, moderate when the percentage is between 1% and 5%, and mild when the percentage is greater than 5% and less than 40% (Srivastava et al., 2013). The hemorrhagic clinical will be present in direct relation to the degree of fault of these plasma factors.

This pathology affects 1 in 5-6,000 live male births in the case of hemophilia A and 1 in 30,000 in the B type and, although prevalence data vary depending on the country, it is estimated at 1 in 12,000 (Stonebraker, Bolton-Maggs, Michael Soucie, Walker, & Brooker, 2010). However, despite this low incidence and prevalence, the cost associated with this pathology is one of the highest in the healthcare system due to the high cost of treatment, emphasizing the substitutive or replacement therapy of the deficient clotting factor (Makris, 2012).

Joint injury due to intra-articular bleeding episodes, known as hemarthrosis, is the most common symptom of bleeding and higher morbidity in these patients (Leslie & Catherine, 2007). Its main features are the joint swelling and iron deposits in the synovial tissue. Without adequate treatment, recurrent hemarthrosis cause hemophilic arthropathy characterized by chronic inflammation of the synovial membrane, cartilage degradation and bone injury (Roosendaal et al., 1998). This arthropathy entails chronic pain and functional disability of the joint (Forsyth et al., 2012). In order to ensure a good quality of life related to health in people with hemophilia, a continuous evaluation and treatment of these lesions by the specialists is needed.

Although classical statistics allows us to analyze the obtained data, there exist other more advanced analysis methods that allow us to obtain information about existing relationships between variables and to establish population characteristics in a visual way. To obtain this information, the use of Self-Organizing Maps (SOM) is proposed in this paper (Kohonen, 1989). These models make possible to analyze all the variables and input patterns together. Furthermore, this analysis is done visually, allowing to draw conclusions about the data intuitively, according to human perception skills (Soukup & Davidson, 2002). The SOM and its variants, are used in all kinds of fields such as finance (Huang & Wu, 2010), marketing (Seret, Verbraken, Versailles, & Baesens, 2012), bioinformatics (Ghouila et al., 2009) and biomedicine (Aoki et al., 2011). In Ref. (Lebbah, Bennani, & Rogovschi, 2008) the authors present a probabilistic self-organizing map for the analysis and visualization of multivariate binary data. Moreover, in Ref. (Zehraoui & Bennani, 2005) a self-organizing maps for sequence clustering and classification is presented. Finally, in Ref. (Horio & Yamakawa, 2001) the authors present a feedback self-organizing map for spatio-temporal pattern classification. To the best of our knowledge, SOM has not been previously used in the analysis of hemarthrosis data in patients with haemophilia. Therefore, this study is aimed to analyze the relationships existing in the clinical and sonographic variables related to hemarthrosis by using SOM.

The remainder of the paper is structured as follows: In Section 2 the used data is described. In Section 3 the basics of self-organizing maps are detailed. In Section 4 the process used to generate the maps and the obtained results are described. Finally, Section 5 summarizes the conclusions of this work.

2. Data

The data set analyzed in this article was obtained through clinical visits to 30 hemophilics patients from the Valencian Community, Spain. A total of 155 visits were carried out during the follow-up of

50 intraarticular bleeding episodes. These patients are usually treated in the Haemostasis and Thrombosis Unit of the University and Polytechnic Hospital La Fe of Valencia. It must be noted that, due to the low incidence of this disease, the number of patients included in this study is a representative portion of the existing hemophilia population in the Valencian Community (366 patients registered: 121 severe, 36 moderate and 209 mild).

In each visit the patient's age, type of hemophilia, explored joint, existence of haemarthrosis, if it was spontaneous or induced, joint swelling, Range of Movement (ROM) of the joint, joint pain, the existence of intra-articular effusion as well as existence of synovitis and its severity were recorded. To measure the swelling, a conventional measuring tape (perimeter of the joint in cm) was used, for the ROM a goniometer and for the pain, a visual Analog Scale (VAS). To diagnose the existence of intra-articular effusion and synovitis, a portable ultrasound machine Mindray DP-6600® (Mindray Bio-Medical Electronics Co. Ltd, Shenzhen, China) with a 7.5 MHz linear array transducer was used. A longitudinal section was made in each joint (Aznar et al., 2015). The "joint swelling" variable is obtained by calculating the difference between the joint perimeter measured in the visit with regard to the perimeter presented in basal, i.e. without injury. Therefore, if the variable value is 0, there is no swelling while positive values indicate the presence of swelling. Similarly, the "ROM" variable is obtained from the difference in degrees between the mobility of the joint measured in the bleeding episode and mobility in base (from the point of maximum extension to maximum flexion). Negative values indicate loss of mobility. Due to the fact that it is a group of patients with a continued medical monitoring, basal data of joint swelling and range of motion of the joints studied for each patient is available. In the case of the qualitative (non-numeric) variables, a codification of its values was performed. Finally, the data set was normalized with mean equal to 0 and variance equal to 1. In Table 1, the names of the variables that were recorded, their description and coding, in the case of qualitative variables, are shown. Moreover, in Table 2 a basic statistics of the numeric variables are presented.

Regarding the hemophilia diagnosis, 29 from the 30 patients are hemophilic type A and 1 patient is hemophilic type B. In Table 3, the number of cases by type of hemophilia, explored joint and synovitis level are shown. In the case of hemophilia type, it is observed that 90% of patients suffer from severe hemophilia, while 10% remaining suffer moderate or mild hemophilia. Furthermore, ankle has been the joint more explored, followed by the knee and finally the elbow. Regarding the synovium alteration, there are 10 cases with level II, 7 with level I, 7 with level III and 6 without synovitis.

Table 1. Variable names, description and codification.

Variable	Description	Codification
Age	Patient's age in years	No cod.
HF	Haemophilia type (severe, moderate, light)	3, 2, 1
Joint	Explored joint (elbow, ankle, knee)	1, 2, 3
Hemarthrosis	Existence of hemarthrosis (yes, no)	1, 0
S/P	Spontaneous or provoked hemarthrosis	1 ,0
Swelling	Perimeter joint swelling (cm)	No cod.
Loss of ROM	Loss of useful range of movement(^{o})	No cod.
Pain	Visual analog scale of pain (0-100)	No cod.
Liquid	Liquid presence in joint (yes, no)	1, 0
Synovitis	Synovitis level (LIII, LII, LI, No)	3, 2, 1, 0

Table 2. Statistics of the numerical variables.

	Age (years	Swellin g(cm)	Loss of ROM	Pain (0-100)
Mea	30.53	0.34	-15.52	13.23
SD	13.89	0.66	25.48	19.66
Max.	66.00	2.50	5.00	83.00
Min.	7.00	-1.00	-120.00	0.00

SD: Standard deviation. ROM: Range of movement.

Table 3. Hemofilia type, explored joint and synovitis level.

Hemophilia type			Explored joint			Synovitis level		
Value	N	%	Value	N	%	Value	N	%
Severe	27	90.0	Elbow	7	23.3	LIII	7	23.3
Moderate	2	6.7	Ankle	12	40.0	LII	10	33.3
Mild	1	3.3	Knee	11	36.7	LI	7	23.3
						No	6	20.0

3. Self-Organising Map (SOM)

SOM is an artificial Neural Network (ANN) developed by Teuvo Kohonen in 1982 (Kohonen, 1982). An updated review can be found in Ref. (Kohonen, 2001). They are very useful in high dimensionality data visualization tasks. SOMs mapped the complex relationships existing between the variables in a high-dimensional data set in a low-dimensional more simple relationship, facilitating the visualization and the information extraction (Haykin & Haykin, 2009).

It is formed by a series of neurons arranged in two layers: a first layer formed by N neurons (one neuron for each input variable) and an output layer in which the information is processed. The second layer is arranged in a regular low-dimensional grid. This layer usually follows a two-dimensional structure, since the information extraction is easier than in a structure with a larger number of dimensions. In the resulting 2-dimensional map or "*components plane*", neighborhood relations existing in the original data space are preserved, so that data of similar characteristics are located closer to each other, whereas those with different characteristics are located furthest (Villmann, Der, Herrmann, & Martinetz, 1994). Therefore, a grouping of data is produced according to its characteristics. The number of neurons required for forming the network varies depending on the number of input patterns, being able to contain from a few dozens to a several hundred. Each neuron is represented by a weight vector *N*-dimensional (prototype vector) defined as $\boldsymbol{m_{ij}} = [m_{ij}^1, \dots, m_{ij}^N]$, where *N* is equal to the number of dimensions of the input vector. Neurons that compose the network are connected to adjacent neurons by a neighborhood relationship. This relationship sets the network structure and may be rectangular, hexagonal, cylindrical or toroidal. The most common structure, and the used in this work, is the rectangular.

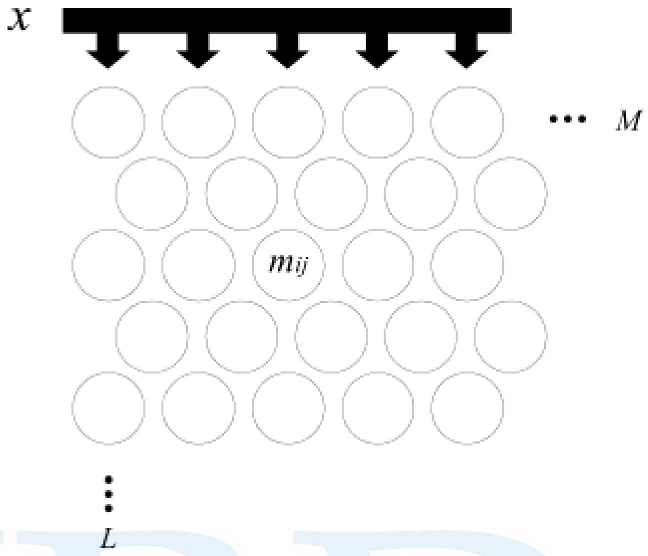

Fig. 1. Array of nodes scheme in a two-dimensional SOM grid.

Before starting training, it is necessary to initialize the weights of the prototype vector of neurons. This initialization can be random or linear type. In the next step, the training algorithm adjusts the weight values by an iterative process. In each iteration a pattern x of the input data is selected and the distance between the selected pattern and the prototype vectors of the map are calculated. The distance measure used is usually Euclidean distance. The neuron whose weight vector is closest to the input pattern x is called *Best-Matching Unit* (BMU) denoted by c:

$$\|\mathbf{x}-\mathbf{m_c}\| = \min_{ij}\{\|\mathbf{x}-\mathbf{m_{ij}}\|\}$$

where ‖ ‖ is the distance used.

This weights adjustment process is similar to the algorithms based on *Vector Quantization* (VQ), as the algorithm of *k-means clustering* (Gray, 1984). The main difference is that in addition to updating the weight vector of the network, the topological neighborhood relations are also updated, placing the

BMU into a closer position to the input vector in the input space. This process is shown in Figure 2. The rule for updating the weight vector of the SOM unit *ij* is defined as:

$$\mathbf{m_{ij}}(t+1) = \mathbf{m_{ij}}(t) + \propto(t) h_{cij}(t)[\mathbf{x}(t) - \mathbf{m_{ij}}(t)],$$

where *t* denotes time. The ***x(t)*** is an input vector randomly selected from the input data set at time *t*, $h_{cij}(t)$ the neighbourhood kernel around the winner unit c and $\alpha(t)$ the learning rate at time *t*. The neighbourhood kernel is a non-increasing function of time and of the distance of unit *ij* from the winner unit *c*. It defines the region of influence that the input sample has in the SOM.

The training is usually performed in two phases. In the first phase, relatively large initial learning rate and neighbourhood radius are used. In the second phase, both learning rate and the neighbourhood radius are small from the beginning. This procedure corresponds to start tuning the SOM approximately to the same space as the input data and then fine-tuning the map in the second phase.

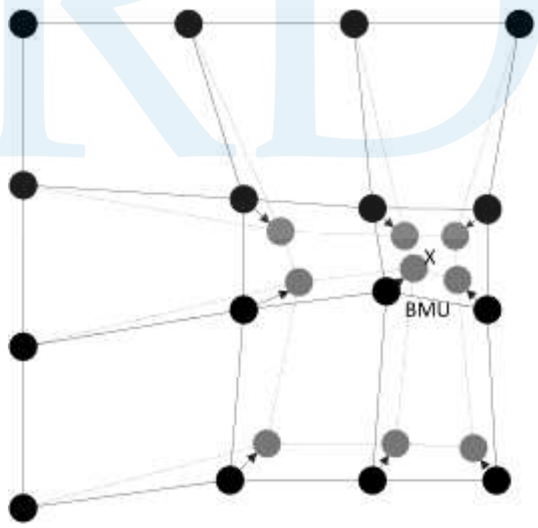

Fig. 2. Updating the BMU and its neighbours towards the imput sample x. The solid and dashed lines correspond to the situation before and after updating, respectively.

4. Results

This section describes the process carried out to tuning the parameters of SOMs (number of neurons, network topology, training algorithm, etc.) as well as the method used to obtain the optimal network. Afterwards the results obtained with data set described in Section 2 are explained in Section 4.2.

4.1. Settings maps

For this work SOM library developed in the *Laboratory of Information and Computer Science in the Helsinki University of Technology* has been used. This library contains all the algorithms needed to train, analyze and visualize SOMs.

First, the number of neurons (map size) was adjusted to adapt it to the number of variables and patterns in the data set. Size was chosen originally as $5\sqrt{n}$, where n is the number of training samples or patterns (Vesanto, Alhoniemi, Himberg, & Kiviluoto, 1999). For the used data set, the number of neurons is $5\sqrt{155} = 62$, so that the SOM algorithm implements a rectangular map of 7 neurons wide by 9 high (63 neurons). In the experiments it was found that the number of neurons was too large, since a large number of neurons were empty, i.e., no input patterns were assigned. Adjusting the number of neurons finally a 5x5 size rectangular map was set. Then, a different training options were tested by combining all possibilities. Specifically, a total of 4008 maps were trained combining the different types of weights initialization vectors of the neurons (random and linear), with the available types of training algorithms (batch and sequential) and with the 4 types of neighborhood functions (gaussian, cut gaussian, epanechicov, bubble). The remaining parameters such as the training time, learning rate and the neighborhood radius, were adjusted according to the recommended values in SOM toolbox documentation (Vesanto et al., 1999).

To select the best network is usually evaluated 2 types of criteria: resolution or quantization error and the topology conservation or topographic errors (Kohonen, 2001). The quantization error is computed as the average of the distance between each data vector and its BMU. The lower the average, the smaller the error, and therefore the smaller the error in the representation of the data by the SOM. Moreover, the topographic error indicates whether neighborhood relations between the original space and the final space is preserved, and it is calculated as the proportion of all of the data vectors for which the first and second BMU are not adjacent units. It was found that the topographic error was 0 in approximately, 9% of the networks trained. Of the networks with the topographical error equal to

0, the network with less quantization error was chosen in order to obtain the network with a lower combined error. The selected network corresponded to a random initialization of weights of the neurons, *batch* training algorithm and *gaussian* neighborhood function.

4.2. Visual data mining

After selecting the optimal network, the obtained results were analyzed. In Figure 3 the "*Hits map"* is shown. In this figure, the rectangular 5 x 5 neurons topology configured for the SOM can be observed. Each hexagon represents a neuron or node. It also shows the network response to the input data so that the colored area within each hexagon is proportional to the number of patterns that is similar to that neuron coefficients. Therefore in the "*Hits map*" the distribution is represented in the network of BMUs corresponding to the input data.

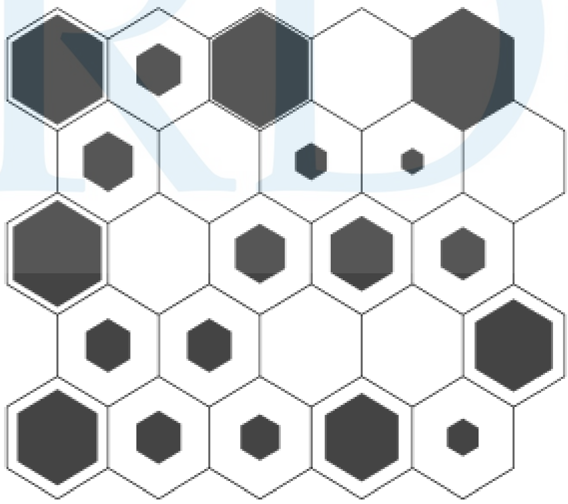

Fig. 3. "*Hits map*" of the trained SOM network. The BMUs and the network response to the input patterns are shown.

In Figure 4, the obtained "*Components plane*" is shown. In this figure, all variables are shown together. Examining and comparing different maps, the relation-ships between variables can be obtained.

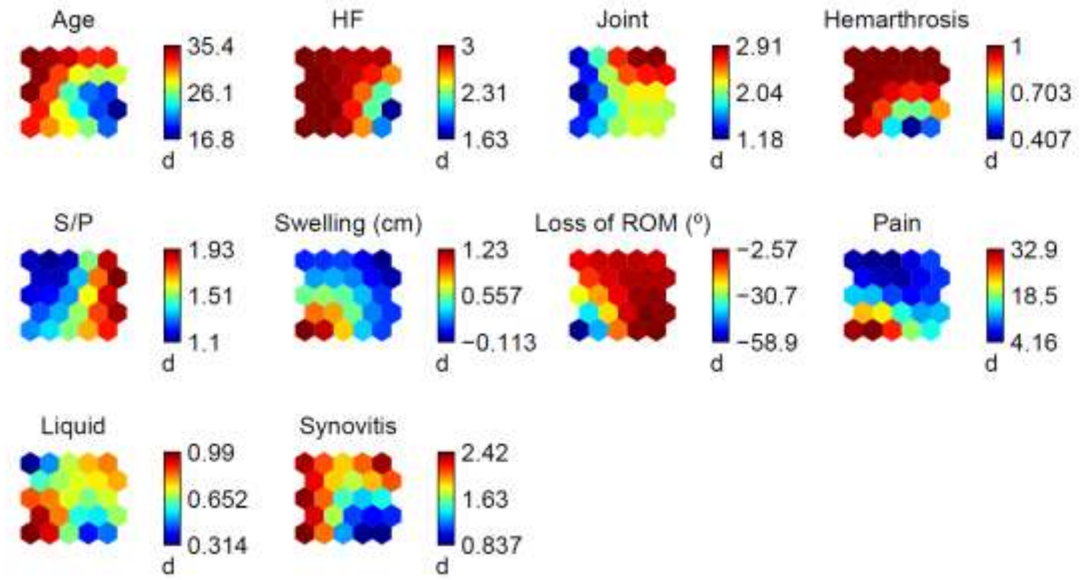

Fig. 4. "*Components plane*" obtained with the optimal SOM network. It allows to extract the relationships between variables.

First, the relationship between loss of ROM and pain is discussed below. In the Loss of ROM (º) component, corresponding to a loss of useful range of movement in joint, the scale is defined between -58.9º (dark blue) and -2.57º degrees (dark red). The majority of the figure nodes appear represented in different shades of red, representing small losses of mobility. However, the nodes located in the lower left corner appear in shades of blue, representing great losses of mobility. Moreover, on the component corresponding to the pain, scale is defined between 4.16 (blue) and 32.9 (red). In this case the blue nodes (low pain) predominate, whereas in the lower left corner, nodes that represent higher levels of pain (different shades of red) are located. Comparing both variables, a clear relationship between the loss of mobility and the level of pain that suffers the patient can be observed. When the loss of mobility is high (blue colors), pain is high (red colors). Additionally, these nodes are located in similar areas of the maps, representing that there is a clear data grouping. Afterwards, the possible relationships among swelling, fluid and sinovitis were examined. In the joint swelling, the scale ranges from -0.11 cm to 1.23 cm approximately. In some patients, the joint contains fluid in basal conditions, so that after treatment, the perimeter joint may be lower than at baseline. This explains the negative values given in the scale. Dividing the figure into 4 parts, we see that, in the bottom left, the data corresponding to high levels of swelling are grouped. In the rest of the map, the swelling is low and it is represented in various shades of blue. On the other hand, in the variable corresponding to the presence of liquid, the nonexistence of liquid is represented as 0 and the cases in which it exists as 1. It is observed that the color groups are mixed and the only clear group is located in the lower left corner. This grouping, composed by different shades of red, symbolizes that the patients in that zone contain fluid in their joints. It is noted that this area corresponds to the area of greatest swelling discussed above. Regarding the level of synovitis component, the clearest group is found in the lower right corner, which represents cases where there is no synovitis (0) or the level is I. Moreover, in the area coinciding with high swelling and existence of liquid, it is observed that the predominant synovitis levels are III and II.

Comparing the remaining variables more relations can be obtained. For example, patients with medium-high ages (red nodes), with severe hemophilia (red nodes), have suffered haemarthrosis (red nodes) mostly in elbows (blue nodes). The hemarthrosis appeared spontaneously (blue nodes), swelling has been high (red nodes) as well as loss of ROM (blue nodes) and pain (red nodes). Furthermore, it has also been observed the existence of joint fluid (red nodes) and the synovitis level has been II and III (red nodes). Focusing on another map area of interest, for example the lower right corner, we found that the younger patients with ages below 26, who suffered haemarthrosis have moderate or mild hemophilia. In these cases, joint bleeds have been fundamentally caused by accident (for example by a blow) and with small severity because the swelling has been slight, the loss of ROM has been small, they have suffered little pain and they have not been suffering synovitis, or this is level I. By looking other maps areas, more conclusions can be obtained. As can be observed, we can extract visually and easily a multitude of relationships between the variables.

5. Conclusions

In this paper it has been proposed the problem of analyzing the relationships between the different variables recorded during the diagnosis and tracking of hemarthrosis in elbows, knees and ankles, that are the most often affected joints in patients with haemophilia. To obtain these relationships, possible groupings of data, as well as relevant information, the use of a visual technique known as Self-Organizing Maps has been proposed. First, a previous analysis of the variables has been performed, coding the qualitative variables. Afterwards, the size and topology of the network have been tuned and the different training parameters, combining all possibilities, have been set. Finally, the results obtained using the network with smaller quantization and topographic errors have been analyzed. The results clearly shown a number of clusters in the data. In particular, is evidenced that patients with more severe lesions appear in a different areas from patients with milder lesions. It has been demonstrated that exists a direct relationship between the presence of fluid in the joint, swelling, loss of ROM, pain and the level of synovitis. In general, patients under 26 who suffered haemarthrosis had mild or moderate hemophilia and the cause of the bleeding was caused (for example by a blow). However, the haemarthrosis severity has been small since they have low values in the studied variables (little swelling, little loss of ROM, low pain, etc.).

Self-organizing maps have proven to be a useful tool to extract visual information from data quickly and easily. It allows to reduce a multidimensional problem in a two-dimensional problem, maintaining the relationships between the variables and, therefore, making easier the knowledge extraction. This is the first time that SOM has been used to analyze haemophilic data. The analysis performed in this study could be extended by adding a greater number of patients, patterns and variables of interest in haemophilic patients and their comorbidities. For example, it would be interesting to relate variables associated with bone mineral density, physical condition, clinical and radiological tests, etc. With this analysis, it would be possible to obtain an overall view of the haemophilia population characteristics.

References

Aoki, S., Hoshi, K., Kawakami, J., Sato, K., Satoh, K., Mori, K., … Yoshida, K. (2011). Assisting the diagnosis of Graves' hyperthyroidism with pattern recognition methods and a set of three routine tests parameters, and their correlations with free T4 levels: Extension to male patients. *Biomedicine & Pharmacotherapy*, *65*(2), 95–104.

Aznar, J. A., Pérez-Alenda, S., Jaca, M., García-Dasí, M., Vila, C., Moret, A., … Bonanad, S. (2015). Home-delivered ultrasound monitoring for home treatment of haemarthrosis in haemophilia A. *Haemophilia*, *21*(2), e147–e150.

Forsyth, A. L., Rivard, G.-é., Valentino, L. A., Zourikian, N., Hoffman, M., Monahan, P. E., … Forriol, F. (2012). Consequences of intra-articular bleeding in haemophilia: science to clinical practice and beyond: CONSEQUENCES OF INTRA-ARTICULAR BLEEDING IN HAEMOPHILIA. *Haemophilia*, *18*, 112–119.

Ghouila, A., Yahia, S. B., Malouche, D., Jmel, H., Laouini, D., Guerfali, F. Z., & Abdelhak, S. (2009). Application of Multi-SOM clustering approach to macrophage gene expression analysis. *Infection, Genetics and Evolution*, *9*(3), 328–336.

Gray, R. (1984). Vector quantization. *IEEE ASSP Magazine*, *1*(2), 4–29.

Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines* (3rd ed). New York: Prentice Hall.

Horio, K., & Yamakawa, T. (2001). FEEDBACK SELF-ORGANIZING MAP AND ITS APPLICATION TO SPATIO-TEMPORAL PATTERN CLASSIFICATION. *International Journal of Computational Intelligence and Applications*, *01*(01), 1–18.

Huang, S.-C., & Wu, T.-K. (2010). Integrating recurrent SOM with wavelet-based kernel partial least square regressions for financial forecasting. *Expert Systems with Applications*, *37*(8), 5698–5705.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69.

Kohonen, T. (1989). *Self-organization and associative memory* (3. ed). Berlin: Springer.

Kohonen, T. (2001). *Self-Organizing Maps*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Lebbah, M., Bennani, Y., & Rogovschi, N. (2008). A PROBABILISTIC SELF-ORGANIZING MAP FOR BINARY DATA TOPOGRAPHIC CLUSTERING. *International Journal of Computational Intelligence and Applications*, *07*(04), 363–383.

Leslie, R., & Catherine, M. (2007). Modern management of haemophilic arthropathy. *British Journal of Haematology*, *136*(6), 777–787.

Makris, M. (2012). Prophylaxis in haemophilia should be life-long. *Blood Transfusion*.

Mannucci, P. M., & Tuddenham, E. G. D. (2001). The Hemophilias — From Royal Genes to Gene Therapy. *New England Journal of Medicine*, *344*(23), 1773–1779.

Roosendaal, G., Vianen, M. E., Wenting, M. J., van Rinsum, A. C., van den Berg, H. M., Lafeber, F. P., & Bijlsma, J. W. (1998). Iron deposits and catabolic properties of synovial tissue from patients with haemophilia. *The Journal of Bone and Joint Surgery. British Volume*, *80*(3), 540–545.

Seret, A., Verbraken, T., Versailles, S., & Baesens, B. (2012). A new SOM-based method for profile generation: Theory and an application in direct marketing. *European Journal of Operational Research*, *220*(1), 199–209.

Soukup, T., & Davidson, I. (2002). *Visual data mining: techniques and tools for data visualization and mining*. New York: Wiley Publ.

Srivastava, A., Brewer, A. K., Mauser-Bunschoten, E. P., Key, N. S., Kitchen, S., Llinas, A., … Treatment Guidelines Working Group The World Federation Of Hemophilia. (2013). Guidelines for the management of hemophilia. *Haemophilia*, *19*(1), e1–e47.

Stonebraker, J. S., Bolton-Maggs, P. H. B., Michael Soucie, J., Walker, I., & Brooker, M. (2010). A study of variations in the reported haemophilia A prevalence around the world. *Haemophilia*, *16*(1), 20–32.

Vesanto, J., Alhoniemi, E., Himberg, J., & Kiviluoto, K. (1999). Self-Organizing Map for Data Mining in MATLAB: The SOM Toolbox. *Simulation News Europe*, (25), 54.

Villmann, T., Der, R., Herrmann, M., & Martinetz, T. (1994). Topology Preservation in Self-Organizing Feature Maps: General Definition and Efficient Measurement. In B. Reusch (Ed.), *Fuzzy Logik* (pp. 159–166). Berlin, Heidelberg: Springer Berlin Heidelberg.

Zehraoui, F., & Bennani, Y. (2005). NEW SELF-ORGANIZING MAPS FOR MULTIVARIATE SEQUENCES PROCESSING. *International Journal of Computational Intelligence and Applications*, *05*(04), 439–456.